

DOCUMENT RESUME

ED 067 393

TM 001 788

AUTHOR Burke, John P.; Elashoff, Janet Dixon  
TITLE The Effects of Serial Dependence on Polynomial  
Regression Models for Individual Growth Data.  
INSTITUTION Stanford Univ., Calif. School of Education.; Stanford  
Univ., Calif. Stanford Center for Research and  
Development in Teaching.  
SPONS AGENCY Office of Education (DHEW), Washington, D.C.  
REPORT NO R-D-MEMO-74  
BUREAU NO BR-5-0252  
PUB DATE May 71  
CONTRACT OEC-6-10-078  
NOTE 43p.  
  
EDRS PRICE MF-\$0.65 HC-\$3.29  
DESCRIPTORS Data Analysis; \*Individual Development; \*Mathematical  
Models; \*Methods Research; Models; \*State of the Art  
Reviews; \*Statistical Analysis  
IDENTIFIERS \*Polynomial Regression Models

ABSTRACT

This paper provides a survey of models for the analysis of individual growth data emphasizing the problems posed by serial or time dependence in the application of polynomial regression models. The concepts of serial correlation and autoregressive models are illustrated. It is demonstrated that standard inference procedures may be quite misleading when applied to polynomial regression models involving time dependence. Little consideration has been given in the literature for the case of individual series to the development of alternative procedures or to the problem of providing a more reliable basis for inference except for the econometric model.  
(Author)

ED 067393

DE-52-505  
TM



Stanford Center for Research and Development in Teaching

SCHOOL OF EDUCATION

STANFORD UNIVERSITY

FILMED FROM BEST AVAILABLE COPY

ED 067393

STANFORD CENTER  
FOR RESEARCH AND DEVELOPMENT  
IN TEACHING

Research and Development Memorandum No. 74

THE EFFECTS OF SERIAL DEPENDENCE ON  
POLYNOMIAL REGRESSION MODELS FOR  
INDIVIDUAL GROWTH DATA

John P. Burke and Janet Dixon Elashoff

School of Education  
Stanford University  
Stanford, California

May 1971

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the Office of Education. (Contract No. OEC-6-10-078, Project No. 5-0252-0704.)

### Introductory Statement

The central mission of the Stanford Center for Research and Development in Teaching is to contribute to the improvement of teaching in American schools. Given the urgency of the times, technological developments, and advances in knowledge from the behavioral sciences about teaching and learning, the Center works on the assumption that a fundamental reformulation of the future role of the teacher will take place. The Center's mission is to specify as clearly, and on as empirical a basis as possible, the direction of that reformulation, to help shape it, to fashion and validate programs for training and retraining teachers in accordance with it, and to develop and test materials and procedures for use in these new training programs.

The Center is at work in three interrelated problem areas: (a) Heuristic Teaching, which aims at promoting self-motivated and sustained inquiry in students, emphasizes affective as well as cognitive processes, and places a high premium upon the uniqueness of each pupil, teacher, and learning situation; (b) The Environment for Teaching, which aims at making schools more flexible so that pupils, teachers, and learning materials can be brought together in ways that take account of their many differences; and (c) Teaching Students from Low-Income Areas, which aims to determine whether more heuristically oriented teachers and more open kinds of schools can and should be developed to improve the education of those currently labeled as the poor and the disadvantaged.

Research and Development Memorandum No. 74, which follows, presents a methodological development generated by the Methodology Unit in answer to problems encountered in the analysis of repeated measurements data. Such data analysis problems pose frequent difficulties in data gathered by Center projects.

## Table of Contents

Section	Page
List of Tables . . . . .	vii
List of Figures . . . . .	ix
Abstract . . . . .	xi
Standard Approach Ignoring Dependence: Polynomial Regression . .	3
Models for Serial Dependence . . . . .	9
Effects of Serial Correlation on Conventional Least Squares	
Procedures . . . . .	16
Bias . . . . .	17
Efficiency . . . . .	17
Validity of Hypothesis Tests and Confidence Intervals . . . .	20
Econometric Studies of the Effects of Serial Correlation . .	24
Tests for Serial Correlation . . . . .	27
Conclusions . . . . .	28
References . . . . .	31
Appendix: Orthogonal Polynomial Coefficients . . . . .	34

# List of Tables

Table No.		Page
1	Hypothetical Data for 10 Time Points on Two Individuals . . . . .	7
2	Least Squares Estimates of Orthogonal Polynomial Coefficients for Data in Table 1 . . . . .	7
3	Relative Efficiency of O.L.S. and G.L.S. Estimators of Polynomial Coefficients Based on 5 Time Points . . . . .	21
4	Probability of Type I Error When Nominal Signifi- cance Level Is 0.05 Under Serial Correlation .	21
5	Ratio of Standard Formula for Var $\hat{\beta}_1$ to Correct Formula for Var $\hat{\beta}_1$ Under the Autoregressive Error Model . . . . .	24

### List of Figures

Figure No.		Page
1	Plots of Hypothetical Data for 10 Time Points on Two Individuals from Table 1 with Fitted Curves Showing Mean, Linear, and Quadratic Components . . . . .	8
2	Data Generated by Linear Regression and Stochas- tic Difference Models . . . . .	11
3	Generated Data and Fitted Regression Lines for Model: $y_t = \beta_0 + \beta_1 t + e_t$ , $e_t = \rho e_{t-1} + u_t$ .	18

### Abstract

This paper provides a survey of models for the analysis of individual growth data emphasizing the problems posed by serial or time dependence in the application of polynomial regression models. The concepts of serial correlation and autoregressive models are illustrated. It is demonstrated that standard inference procedures may be quite misleading when applied to polynomial regression models involving time dependence. Little consideration has been given in the literature for the case of individual series to the development of alternative procedures or to the problem of providing a more reliable basis for inference except for the econometric model.

THE EFFECTS OF SERIAL DEPENDENCE ON POLYNOMIAL  
REGRESSION MODELS FOR INDIVIDUAL GROWTH DATA

John P. Burke and Janet Dixon Elashoff<sup>1</sup>

This paper surveys statistical models for the analysis of individual growth data with the major emphasis on the problems posed by serial or time dependence in the application of polynomial regression models. Time is considered to be an important variable, in contrast to situations in which repeated measurements are a device for reducing error variance or a convenience in data collection.

The problems considered in the literature can be distinguished on the number of individuals considered,  $n$ , and the number of measurements per individual,  $p$ . For  $p = 2$  and  $n$  sufficiently large (say, 10 or more), the problem is one of measuring or contrasting group "growth" or "change." An extensive literature in educational and psychological research has been devoted to the analysis of such two-observation repeated measurements data (Cronbach & Furby, 1970; Lord, 1956, 1957, 1958, 1963; McNemar, 1958; Werts & Linn, 1970, and many others). With a larger number of time points, different methods of characterizing change in group data arise. Under certain assumptions about the structure of the data, analysis of variance techniques may be applied to the analysis of regression models with the time measure (generally in orthogonal polynomial form) as dependent variable. Winer (1962, Ch. 7) discusses the simplest form of analysis; Gaito and Wiley (1963) and Bock (1963) discuss more

---

<sup>1</sup>John P. Burke was a Research Assistant at the Center when this paper was prepared; Janet Dixon Elashoff is Coordinator of the Methodology Unit and Assistant Professor of Education at Stanford University.

general approaches and provide an introduction to the biometric literature in which most attention has been given to this situation. Rao (1965) and Grizzle and Allen (1969) have made more recent contributions.

For a single individual or system and  $p$  very large (over 100 ), an extensive literature on the spectral analysis of time series has evolved, primarily in the context of electrical engineering applications (see, e.g., Parzen, 1961). Holtzman (1963) discusses stochastic difference models for psychological data.

The focus here is on applications to data on a single individual in which  $p$  is moderate, say, in the range 5 to 15 . Problems in the investigation of such data include postulating a model to account for the data, estimating parameters of the model to characterize the individual, and testing hypotheses about an individual's curve. Questions of interest may be: What is an individual's average score? Is there growth or learning (a trend over time)? Is that trend linear, quadratic, exponential?

An example of the determination of characteristics of individual learning curves is provided by a study by Stake (1961) in which a theoretical learning function of hyperbolic form was fitted to data obtained over a series of trials, and the parameters of the function used as measures for an individual in subsequent analyses. Stake's procedures, however, take no account of the issues of dependence to be considered in this paper. (In fact, the psychological literature on learning curves generally assumes a nonstatistical error-free model. See, e.g., Estes, 1956.) Other models for individual change over time which do account for probable dependence are entering the literature of education and psychology as more attention

is given to the idea of experimental time series (Gottman, McFall, & Barnett, 1969; McGuire & Glass, 1967).

The following sections (a) review a standard approach to the analysis of individual growth data, based on a polynomial regression model, which ignores possible dependence among the observations; (b) introduce some possible models to account for serial dependency among the observations; (c) discuss the effects of serial dependence on the standard procedures; (d) outline some methods for detecting the existence of serial dependence; and (e) discuss some alternative approaches to the problem.

#### Standard Approach Ignoring Dependence: Polynomial Regression

Suppose that observations,  $y_t$ , of some variable are taken on an individual at  $p$  points in time. A natural and simple model to describe the relationship between  $y_t$  and time,  $t$ , is a polynomial regression model:

$$(1) \quad y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^{k-1} + e_t \quad k \leq p.$$

With this polynomial regression model and some assumptions about  $e_t$ , one can describe an individual's growth curve, and test hypotheses about the initial score for an individual, or about the existence of trends in scores over time. A polynomial regression may be intended as an exact description of the process generating the data, or as an adequate approximation to a more complex model for the relationship between  $y$  and the time measure. Or, one might apply regression techniques, without concern for any underlying process, because the coefficients, the  $\beta$ 's, provide

useful descriptions of patterns in the data (see Fig. 1). Here attention will be restricted to polynomial regression models although many other types of models are possible (see Anderson, in press).

The standard least squares procedure for estimation and tests of hypotheses of the parameters  $\beta_0, \dots, \beta_k$  will be valid if

$$e_t \sim N(0, \sigma_e^2)$$

(2)

$$E(e_t e_{t'}) = 0 \quad t' \neq t.$$

That is, for any  $t$ , the error term  $e_t$  has a normal distribution with mean zero and variance  $\sigma_e^2$ . In addition,  $e_t$  is independent of the error term at any other time  $t'$ ,  $t' \neq t$ . In other words, standard least squares procedures will be valid if the relationship between  $y_t$  and  $y_{t'}$ ,  $t' < t$ , is due solely to the relationship of the means of  $y_t$  and  $y_{t'}$  to  $t$  and  $t'$  and not to any dependence between the value of the observation  $y_t$  and the actual value observed for  $y_{t'}$ .

It is convenient although not essential to assume that the time points are equally spaced, and it is so assumed here. Minor differences in procedure arise for unequally spaced time points in the determination of orthogonal polynomial coefficients (see Appendix). Common practice varies in the specification of the time measure; the  $p$  points being denoted  $0, 1, 2, \dots, p-1$  or  $1, 2, 3, \dots, p$ . The same form of analysis applies to either convention, although the interpretation of regression coefficients may differ.

The  $p$  equations represented by the general equation (1) may be put in matrix form as

$$(3) \quad \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 1 + \beta_2 1 + \dots + \beta_k 1^{k-1} \\ \beta_0 + \beta_1 2 + \beta_2 2^2 + \dots + \beta_k 2^{k-1} \\ \vdots \\ \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^{k-1} \\ \vdots \\ \beta_0 + \beta_1 p + \beta_2 p^2 + \dots + \beta_k p^{k-1} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_t \\ \vdots \\ e_p \end{bmatrix}$$

or 
$$\underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{e}$$

where  $\underset{\sim}{y}$  is the  $p \times 1$  vector of observations,  $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$ ,

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 2^{k-1} \\ 1 & 3 & 9 & 3^{k-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t & t^2 & t^{k-1} \\ 1 & p & p^2 & p^{k-1} \end{bmatrix},$$

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix}, \quad \underset{\sim}{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix}.$$

Assumptions (2) are that  $\underset{\sim}{e}$  has a multivariate normal distribution with

$$\underset{\sim}{E}(\underset{\sim}{e}) = \underset{\sim}{0} ,$$

and variance-covariance matrix

$$\underset{\sim}{\Sigma} = \begin{bmatrix} E(e_1^2) & E(e_1 e_2) & \dots & E(e_1 e_p) \\ E(e_2 e_1) & E(e_2^2) & & \\ \vdots & & \ddots & \\ E(e_p e_1) & \dots & \dots & E(e_p^2) \end{bmatrix} = \begin{bmatrix} \sigma_e^2 & 0 & \dots & 0 \\ 0 & \sigma_e^2 & \dots & 0 \\ & & \ddots & \\ 0 & & \dots & \sigma_e^2 \end{bmatrix} = \sigma_e^2 \underset{\sim}{I} .$$

From least squares theory the estimators of the coefficients are obtained by solving the equation

$$(4) \quad \underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{y} .$$

Under (2) the estimators  $\hat{\beta}_0, \dots, \hat{\beta}_{k-1}$  have a multivariate normal distribution with means

$$E(\hat{\beta}_i) = \beta_i \quad i = 0, \dots, k-1$$

(5) and variance-covariance matrix

$$\underset{\sim}{V}(\underset{\sim}{\hat{\beta}}) = \sigma_e^2 (\underset{\sim}{X}'\underset{\sim}{X})^{-1} .$$

The following hypothetical example illustrates the application of regression techniques to growth-type data. Suppose there are measures at each of 10 equidistant time points on two individuals (Table 1 and Fig. 1). One can fit regression curves and consider how the parameters of these curves reflect apparent differences in the pattern of growth.

In this case it seems appropriate to fit a quadratic curve to both

sets of data. The quadratic model may be expressed with either powers of  $t$  or the orthogonal polynomial forms of these powers as the independent variables (see Appendix). The orthogonal polynomial model for a quadratic is written as

$$y_t = \gamma_0 + \gamma_1 p_{1t} + \gamma_2 p_{2t} + e_t$$

where  $p_{it}$  are the orthogonal polynomial coefficients of degree  $i$ .

Applying standard least squares procedures for orthogonal polynomials (Appendix formula A.4), the data in Table 1 yields the estimates shown in Table 2.

TABLE 1

Hypothetical Data for 10 Time Points on Two Individuals

Individual	Time									
	0	1	2	3	4	5	6	7	8	9
1	20	18	25	22	28	36	50	55	70	73
2	20	33	48	52	66	63	76	78	83	81

TABLE 2

Least Squares Estimates of Orthogonal Polynomial

Coefficients for Data in Table 1

Individual	Coefficients		
	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
1	39.7	3.28	1.38
2	60.0	3.39	-1.42

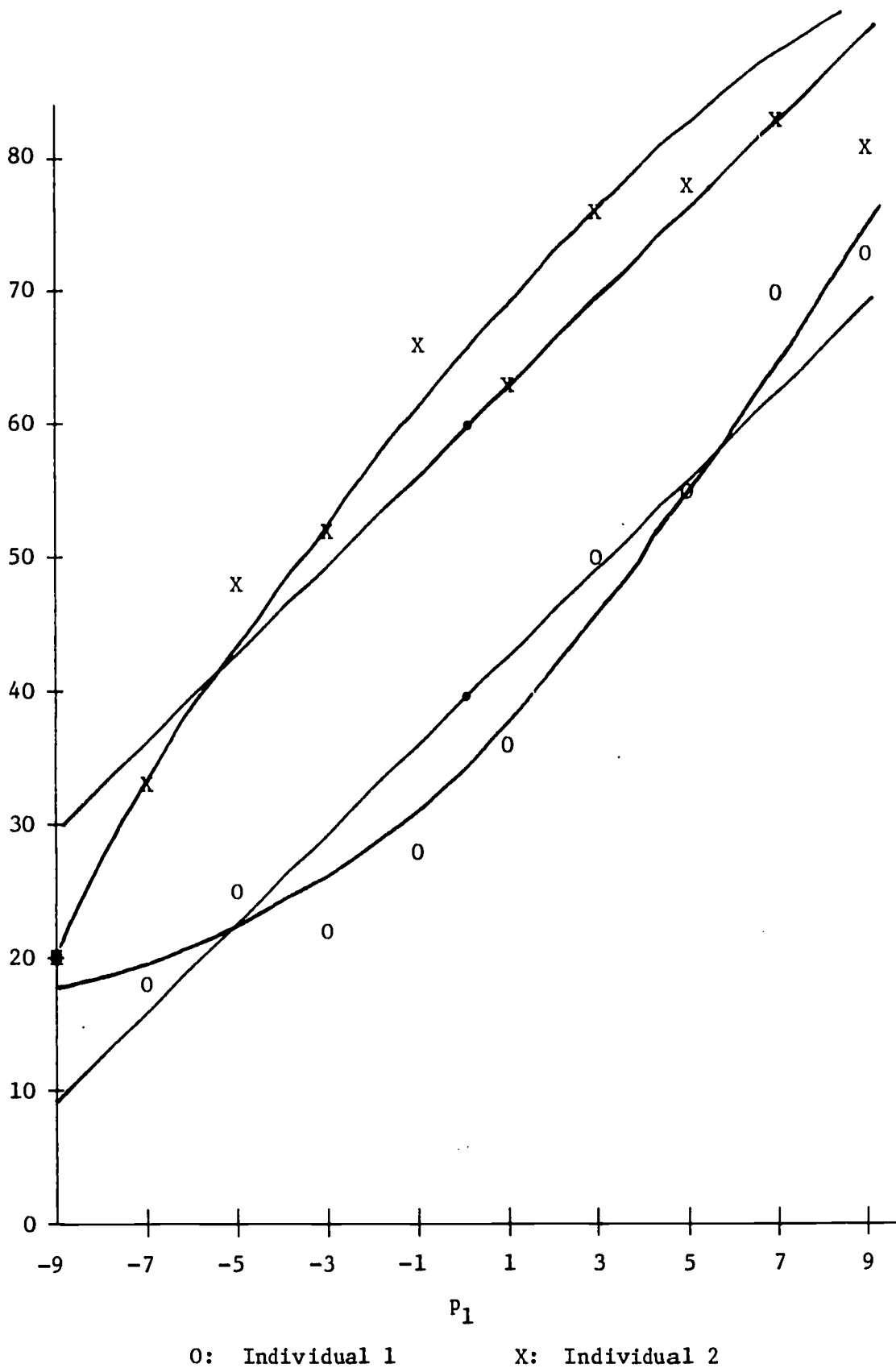


Fig. 1. Plots of hypothetical data for 10 time points on two individuals from Table 1 with fitted curves showing mean, linear, and quadratic components.

Examination of the coefficients shows that the mean  $\hat{\gamma}_0$  (shown by • on Fig. 1) is considerably lower for individual 1. The coefficient  $\hat{\gamma}_1$  which reflects the linear component of the trend (the straight lines in Fig. 1 show mean + linear component) is nearly the same for individuals 1 and 2. The coefficients  $\hat{\gamma}_2$  clearly reflect the different characteristics of the quadratic components of the two individuals. Note that  $\hat{\gamma}_2$  for individual 1 is positive indicating a concave curve, which might be interpreted as "early learning," while  $\hat{\gamma}_2$  for individual 2 is negative indicating a convex curve or "late learning." The curved lines in Figure 1 are the fitted quadratic curves; the difference between the corresponding straight and curved lines being the quadratic component contributed by  $\hat{\gamma}_2$ .

#### Models for Serial Dependence

Suppose a polynomial regression model such as that given by conditions (1) and (2) while being of the right general form does not fit the data well; that is, the value of  $y_t$  seems to have some dependence on the actual value of  $y_{t-1}$ , not explained by the dependence of the mean of  $y_t$  on  $t$ . A number of different models to describe this sort of situation have been proposed.

Models based on stochastic difference equations express the dependence of observations on preceding observations rather than on the time measure. A simple model of this type is the first-order model:

$$(6) \quad y_t = \rho_1 y_{t-1} + e_t,$$

where the error component  $e_t$  is assumed to have the following properties:

$$E(e_t) = 0$$

$$E(e_t e_{t'}) = \begin{cases} \sigma_e^2 & t = t' \\ 0 & t \neq t' \end{cases}$$

$$E(e_t y_{t-s}) = 0 \quad 1 \leq s \leq t-1 .$$

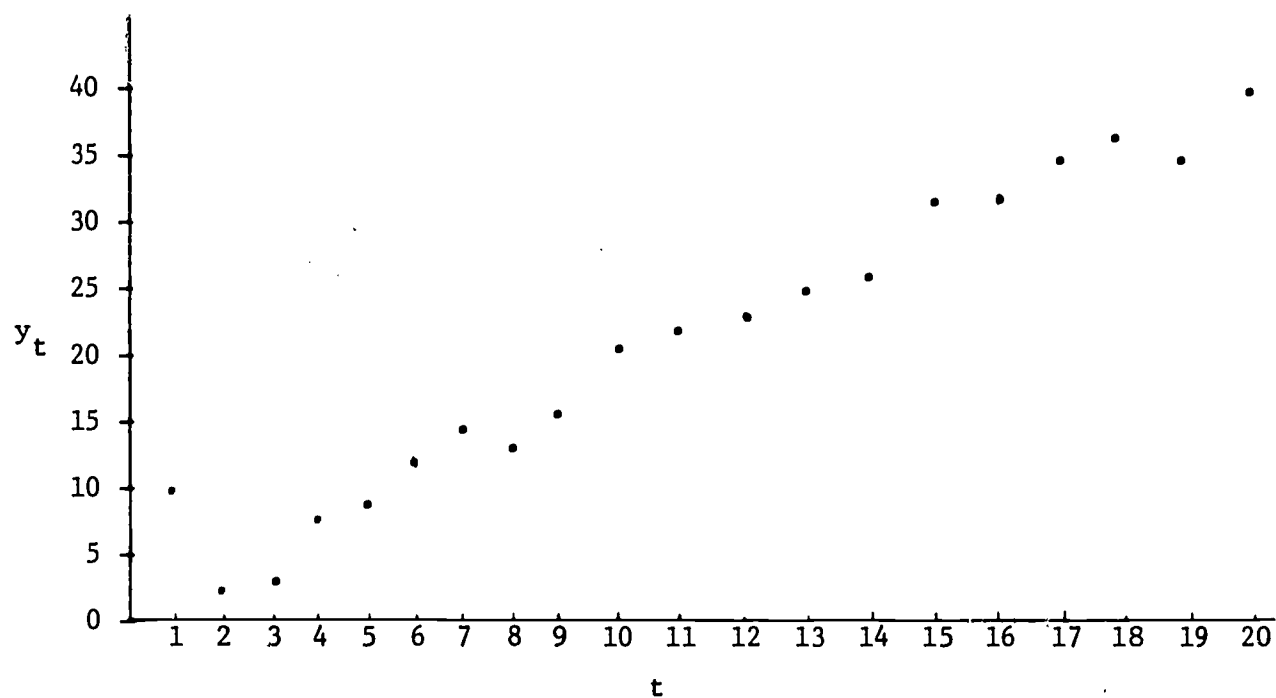
That is, the  $e_t$  have identical distributions with zero mean, constant variance ( $\sigma_e^2$ ), and are independent of the preceding observations.

This model is referred to as a first-order stochastic difference equation, Markov chain, or autoregressive process. In this first-order model an observation is assumed to be dependent on the immediately preceding observation  $y_{t-1}$ , but not directly dependent on any observations preceding  $y_{t-1}$ . Other models may be proposed reflecting higher order dependence such as a stochastic difference equation of the second order

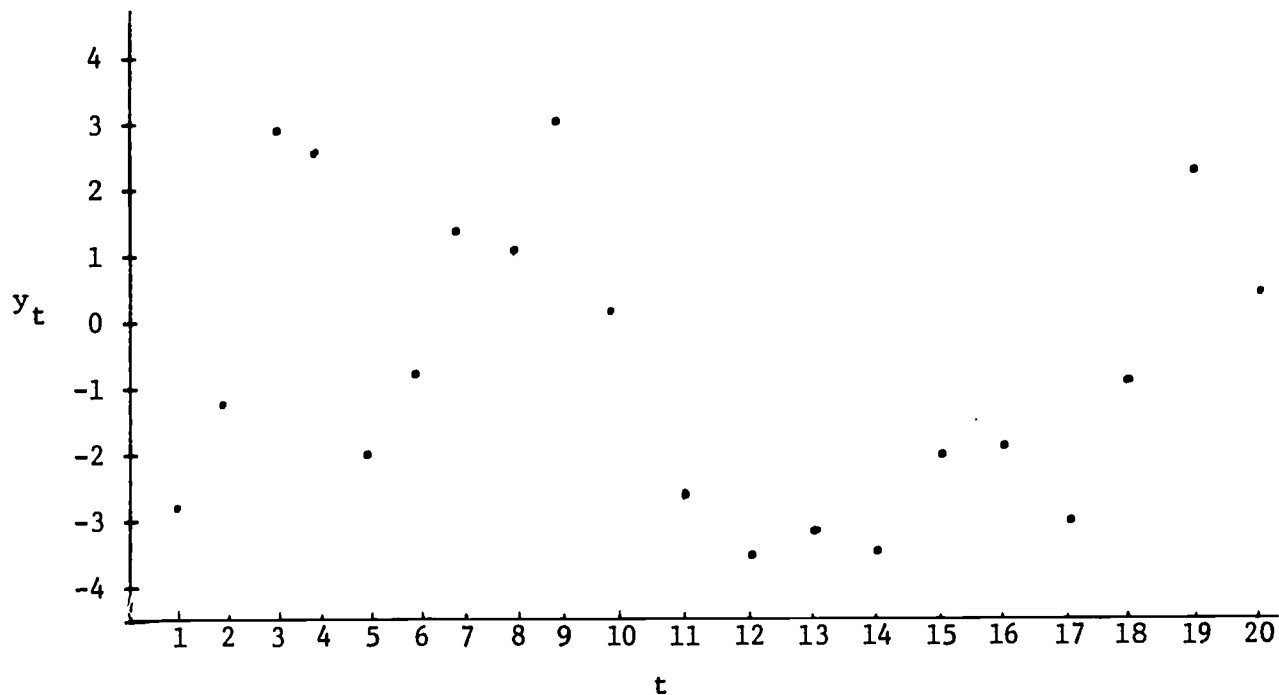
$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + e_t .$$

The graphs of Figure 2 illustrate the kinds of series generated by stochastic difference and regression models. The first-order stochastic difference model generates a series in which there is oscillation about a mean value (0) with no overall trend, and a tendency for runs of positive and negative values. The linear regression model gives rise to a trend and is more appropriate to processes in which an increase in level with time is expected.

A more interesting approach to growth data of this nature may be to retain the basic polynomial regression model described in (1) but to specify a model for serial dependence among the errors or residuals.



Linear Regression Model:  $y_t = 0.4t + e_t$ ,  $e_t \sim N(0,4)$



Stochastic Difference Model:  $y_t = 0.4y_{t-1} + e_t$ ,  $e_t \sim N(0,4)$

Fig. 2. Data generated by linear regression and stochastic difference models (note different scales).

Descriptions of several of the commonly used models of this type will be helpful; each requires different assumptions about the nature of serial dependence among the errors.

A basic assumption of the standard least squares estimates for model (1) is that the errors have constant variance and are independent of each other; i.e., the variance-covariance matrix is  $\Sigma = \sigma_e^2 I$ . Serial correlation or dependence is said to exist whenever there is correlation between any pair of error terms; i.e.,

$$E(e_t e_{t'}) \neq 0 \quad \text{for some } t \text{ and } t' \neq t.$$

Usually a particular model based on empirical evidence or theoretical considerations is chosen to represent the form of serial correlation. Here the characteristics of two particular models will be considered, the autoregressive error model, and the cumulative error model.

The most commonly used model in the literature is the first-order stochastic difference equation or autoregressive error model<sup>2</sup>:

$$(8) \quad e_t = \rho e_{t-1} + u_t,$$

with  $|\rho| < 1$ , and

$$(9) \quad E(u_t) = 0, \quad E(e_t, u_t) = 0 \quad t' < t, \quad E(u_t u_{t'}) = \begin{cases} \sigma_u^2 & t = t' \\ 0 & t \neq t' \end{cases}.$$

$$\text{Hence} \quad E(e_t) = 0 \quad \text{and} \quad E(e_t^2) = \sigma_e^2 = \sigma_u^2 / (1 - \rho^2).$$

In other words, the error term  $e_t$  is a linear function of the error term at time  $t-1$ , plus term  $u_t$  representing additional unaccounted-for variability.

<sup>2</sup>This model for the error term is exactly like the stochastic difference model (6) for observations previously considered.

It can be shown that  $E(e_t e_{t+s}) = \rho^s \sigma_e^2$ . Hence the variance-covariance matrix of errors generated by this process is

$$\Sigma_a = \sigma_e^2 \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{p-1} \\ \rho & 1 & \rho & . & . & \rho^{p-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho^{p-1} & . & . & . & \rho & 1 \end{bmatrix} .$$

Such a pattern of variance-covariance matrix, in which off diagonal correlations decrease monotonically, is often referred to as a simplex.

In the context of this autoregressive model for serial dependence, a population serial correlation coefficient of lag  $s$  is defined as the correlation between error terms  $s$  units apart. Algebraically,

$$(10) \quad \rho_s = \frac{E(e_t e_{t+s})}{\sqrt{E(e_t^2)E(e_{t+s}^2)}} .$$

$$\text{Under model (8), } \rho_s = \frac{E(e_t e_{t+s})}{\sigma_e^2} = \rho^s .$$

The sample serial correlation coefficient of lag 1 of the residuals from regression provides an estimator of  $\rho$ . If one defines the residuals  $z_t = y_t - \hat{y}_t$ , where  $\hat{y}_t$  is the standard least squares estimate at time  $t$ , then one definition of the sample coefficient of lag  $s$  is

$$(11) \quad r_s = \frac{\sum_{t=s+1}^p z_t z_{t-s}}{\sum_{t=1}^p z_t^2}$$

which may be likened to a product moment correlation between the series  $z_{s+1}, \dots, z_p$  and  $z_1, \dots, z_{p-s}$ . This coefficient provides a statistic useful for detecting dependence in observed data.

The cumulative error model is less commonly encountered, but provides an interesting comparison with the previous model. Here it is proposed that

$$(12) \quad e_t = e_{t-1} + u_t,$$

$$\text{or} \quad e_t = \sum_{t=1}^p u_t,$$

and conditions (9) hold for  $e_t$  and  $u_t$ . It follows that  $E(e_t) = 0$ ,

and  $E(e_t e_{t+s}) = t \sigma_u^2$  ( $s \geq 0$ ), and hence the variance-covariance matrix of errors is

$$\Sigma_c = \sigma_u^2 \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & 2 & 2 & \dots & 2 & 2 \\ 1 & 2 & 3 & \dots & 3 & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & 3 & \dots & p-1 & p-1 \\ 1 & 2 & 3 & \dots & p-1 & p \end{bmatrix}.$$

In the cumulative error model, the error at time  $t$  is composed of the error at time  $t-1$  plus an independent component  $u_t$ . The increment  $u_t$  may be conceived of as an error arising during the time interval  $t-1$  to  $t$ . The conditions under which one might expect this model to hold are discussed by Mandel (1957).

Note that the cumulative error model is the same form as the autoregressive model but with  $\rho = 1$ . The class of first-order autoregressive models with  $|\rho| \geq 1$  is often denoted as the class of nonstationary autoregressive processes, while  $|\rho| < 1$  defines the class of stationary autoregressive processes. Stationary processes have constant variance across time; note the ones in the diagonals of  $\Sigma_a$ . The variances of nonstationary processes increase or "explode" over time; note the diagonal terms in  $\Sigma_c$ . Observe that in the consideration of data for a single individual, an estimate of  $\Sigma_a$  or  $\Sigma_c$  would not be available to provide a suggestion of which underlying error model was appropriate. In order to get an estimate of  $\Sigma$ , replications of the sequence of  $p$  observations would be needed.

The cumulative error model generates different between-times correlations as well as different variances. Thus

$$\rho_{t,t+s} = \frac{\sigma_{t,t+s}}{\sigma_t \sigma_{t+s}} = \frac{t \sigma_u^2}{\sqrt{t \sigma_u^2} \sqrt{(t+s) \sigma_u^2}} = \sqrt{\frac{t}{t+s}}.$$

For any value of  $s$  this correlation is a function of  $t$ . Hence there is no single population serial correlation coefficient of lag  $s$ ,  $\rho_s$ , as in the autoregressive model.

The point here is that, while the autoregressive model is the one most frequently encountered in the literature and is often considered to be the model for serial correlation, other models may be worth considering. The cumulative error model is one example. Other possibilities are second-order autoregressive models, or a model such as that used by Box<sup>3</sup>. These models may generate rather different patterns of serial dependence from those generated by the stationary autoregressive model. Furthermore, such concepts as the serial correlation coefficient may not be generally meaningful in other models.

#### Effects of Serial Correlation on Conventional Least Squares Procedures

This section will consider the effects of serial correlation on ordinary least squares procedures for estimation of parameters in (1) in terms of: (a) bias in the least squares estimators of regression coefficients; (b) the efficiency of the least squares estimators; (c) the validity of hypothesis tests and confidence intervals based on the conventional procedure.

The problem of prediction of values beyond the range of the time measure is considered by Johnston (1963, pp. 195-199), in the context of the general regression model encountered in econometric applications.

---

<sup>3</sup>Box (1954), in studying the effects of serial correlation on analysis of variance, used

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & 0 & \dots & 0 \\ \rho & 1 & \rho & \dots & 0 \\ 0 & 0 & 0 & \dots & \rho & 1 \end{bmatrix} .$$

Generally speaking, the results presented in this section may be summarized by saying that the ordinary least squares estimators of the regression coefficients,  $\hat{\beta}$ , are unbiased and reasonably efficient (that is, the variances are not appreciably larger than the minimum obtainable variances) but tests and confidence intervals for the true values of  $\beta$  may be seriously misleading.

Figure 3 illustrates the effect of serial correlation, showing data generated by a linear regression model with autoregressive error term for  $\rho = 0.2$  and  $0.8$ .  $\hat{\beta}_0$  and  $\hat{\beta}_1$  are the ordinary least squares estimates,  $\hat{\rho}$  is the  $r_1$  of (11), and  $d$  is the Durbin-Watson d-statistic (22) whose use will be discussed later. Discrepancies between the line specified by the model and the line fitted using ordinary least squares techniques are evident, particularly for  $\rho = 0.8$ . Note that the residuals from the fitted line tend to be more random than the actual errors. Consequently,  $\hat{\rho}$  is an underestimate of  $\rho$ .

### Bias

It is readily shown that the estimators obtained by ordinary least squares techniques are unbiased under any form of serial correlation (Johnston, p. 188) as long as  $E(e_t) = 0$ . This means that the expected value of a coefficient estimated from replications of a given series is the population coefficient.

### Efficiency

The most efficient estimator of a parameter (out of a class of possible estimators) is that for which the variance of the estimate is least. Although two estimators may both be unbiased, the more efficient

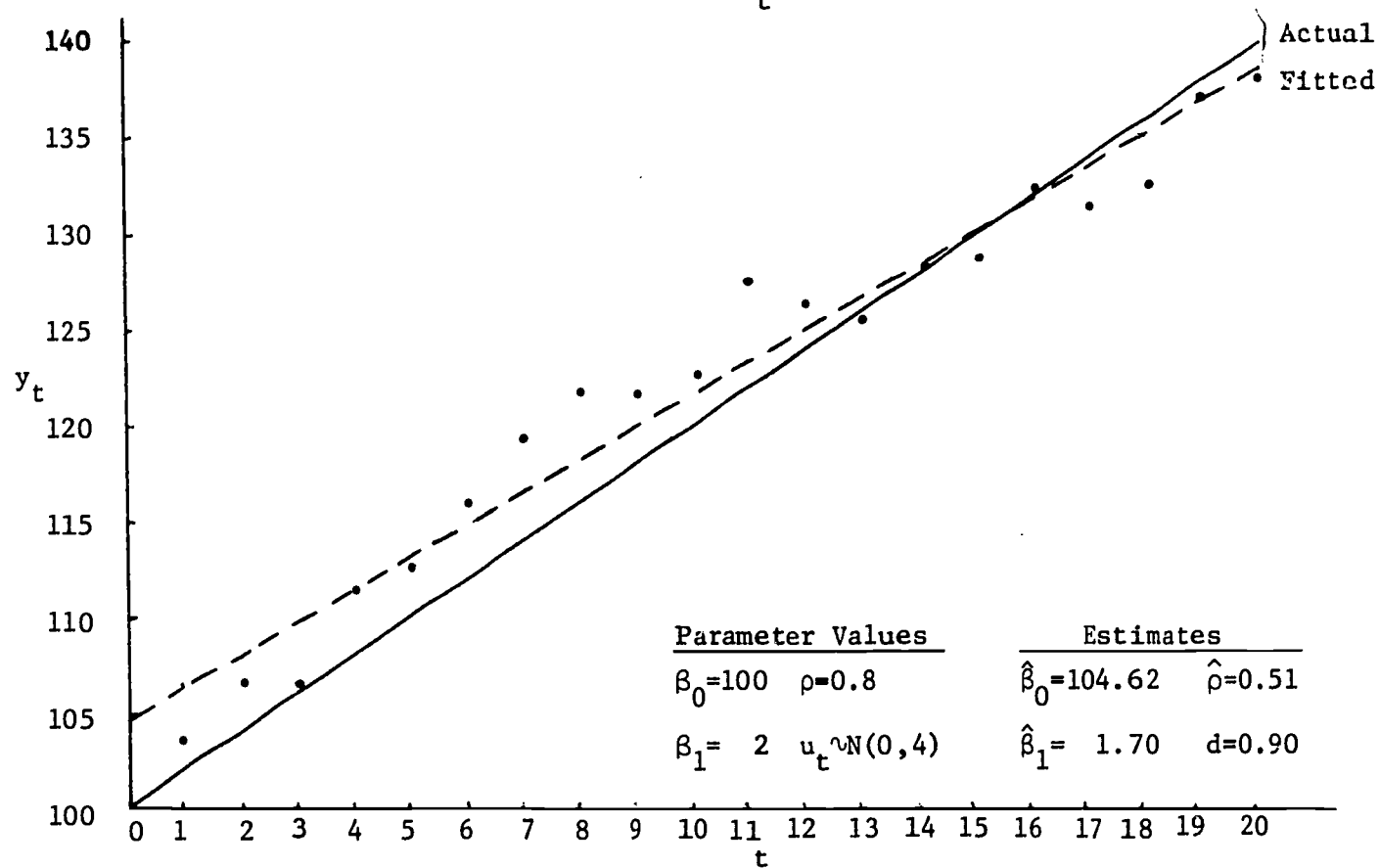
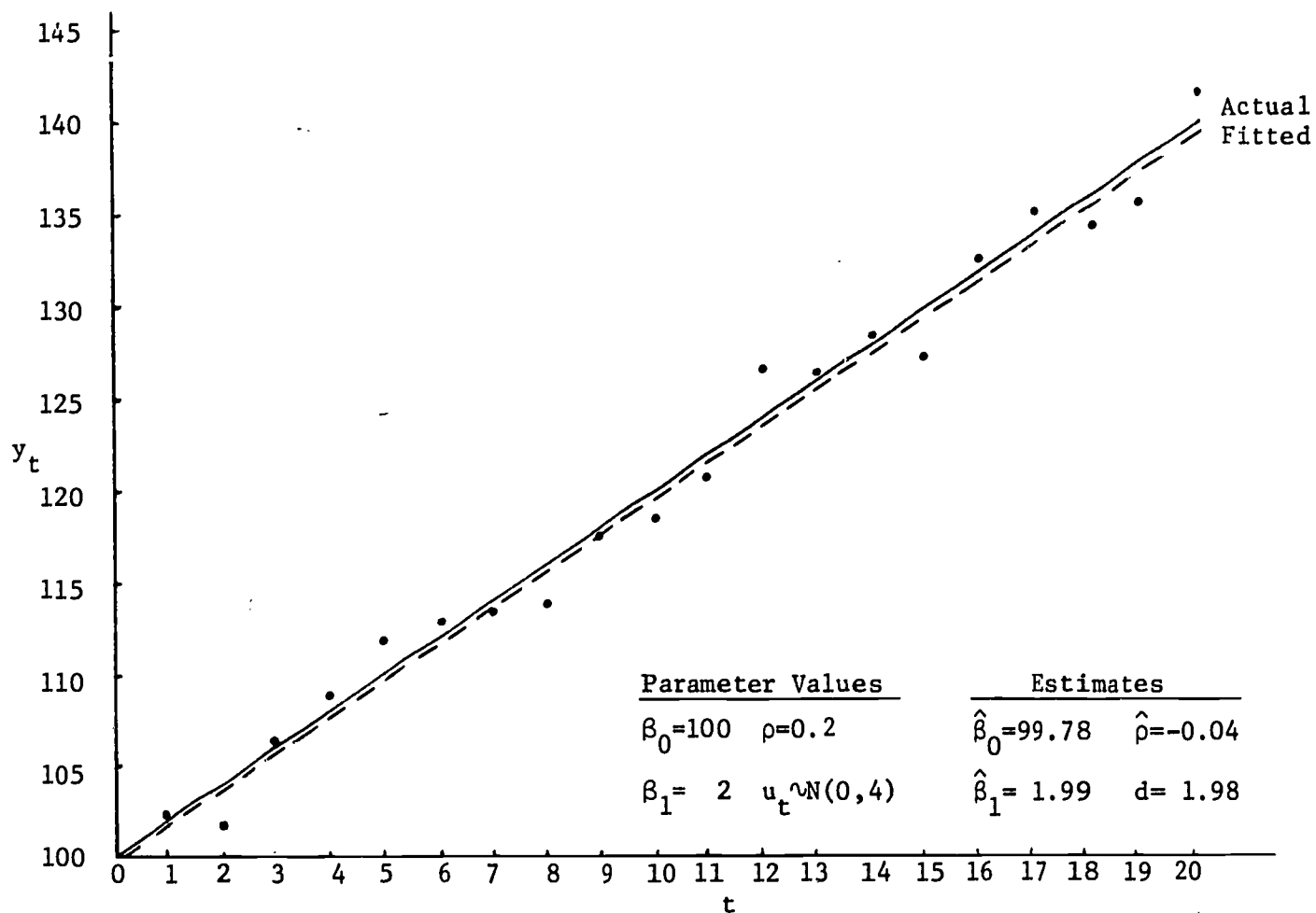


Fig. 3. Generated data and fitted regression lines for model:  
 $y_t = \beta_0 + \beta_1 t + e_t$ ,  $e_t = \rho e_{t-1} + u_t$ .

will tend to provide individual estimates closer to the population value, narrower confidence intervals, and consequently more powerful tests of hypotheses. When the conventional assumptions of independence (2) are met, the ordinary least squares (o.l.s.) estimators of the regression coefficients are the most efficient of any unbiased estimators which are linear functions of the observed values.

When serial correlation exists the least squares estimators may not be the most efficient. If the error variance-covariance matrix  $\Sigma$  is assumed known, it can be shown that the most efficient linear unbiased estimators of the  $\beta$  coefficients are those obtained by techniques known as generalized least squares (Johnston, sect. 7.3). These estimators, the g.l.s. estimators, are given by

$$(13) \quad \underset{\sim}{\beta}^* = (\underset{\sim}{X}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{Y}.$$

In practice,  $\Sigma$  would not be known. However, calculation of  $\frac{\text{Var } (\beta_1^*)}{\text{Var } (\hat{\beta}_1)}$  for various specifications of  $\rho$  and  $\Sigma$  provides a lower

bound for the efficiency of the o.l.s. estimators relative to any other linear unbiased estimators one might devise.

For general error variance-covariance matrix  $\Sigma$ , the variances for g.l.s. and o.l.s. estimators are given by:

$$(14) \quad \underset{\sim}{\text{Var}} (\underset{\sim}{\beta}^*) = (\underset{\sim}{X}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{X})^{-1}$$

$$\underset{\sim}{\text{Var}} (\underset{\sim}{\hat{\beta}}) = (\underset{\sim}{X}' \underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{\Sigma} \underset{\sim}{X} (\underset{\sim}{X}' \underset{\sim}{X})^{-1}.$$

(The formula for  $\text{Var } \hat{\beta}$  reduces to formula (5) when  $\Sigma = \sigma_e^2 I$ , the case of independence.)

Table 3 gives the relative efficiencies of the o.l.s. and g.l.s. estimators of the polynomial coefficients  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  based on five time points, for the autoregressive model ( $\Sigma = \Sigma_a$ ) with  $\rho$  varying from -0.8 to 0.8, and for the cumulative error model. Note that relative efficiency increases with increase in number of time points. (See Rosenblatt, 1956, for tables of variances for o.l.s. and g.l.s. or Markov estimates for 10, 15, 20, and 50 time points in a linear regression model.)

It is clear that the relative efficiency of the o.l.s. estimators is generally high except for large negative  $\rho$ . Decisions between alternative estimators based on relative efficiency depend on the complexity of calculation for the alternative estimators, and problem-specific decisions about power, etc.; however, except for  $\rho < -0.6$ , the o.l.s. estimators have satisfactorily small variances under the autoregressive model.

#### Validity of Hypothesis Tests and Confidence Intervals

There are a variety of hypotheses about the values of regression coefficients that may be tested; e.g.,  $H_0: \beta_1 = 0$  or  $H_0: \beta_0 = \beta_1 = \beta_2 = 0$  ( $\beta = 0$ ). Under the ordinary assumptions of independence (2), t- and F-tests are appropriate for these hypotheses. There is evidence, however, that the existence of serial correlation seriously affects the validity of the tests.

TABLE 3  
Relative Efficiency of O.L.S. and G.L.S. Estimators of  
Polynomial Coefficients Based on 5 Time Points

	<u>Autoregressive Error Model</u>									<u>Cumulative</u>
					$\rho$					<u>Error</u>
	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	<u>Model</u>
<u>Var (<math>\beta_0^*</math>)</u>										
Var ( $\hat{\beta}_0$ )	0.43	0.75	0.92	0.99	1.00	1.00	0.98	0.98	0.96	0.96
<u>Var (<math>\beta_1^*</math>)</u>										
Var ( $\hat{\beta}_1$ )	0.38	0.73	0.92	0.99	1.00	0.99	0.99	0.98	0.98	0.98
<u>Var (<math>\beta_2^*</math>)</u>										
Var ( $\hat{\beta}_2$ )	0.37	0.72	0.92	0.99	1.00	0.99	0.99	0.98	0.98	0.98

TABLE 4  
Probability of Type I Error When Nominal Significance  
Level Is 0.05 Under Serial Correlation

	$\rho$						
	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
Example 1	.0026	.0164	.05	.11	.20	-	-
Example 2	-	-	.05	.14	.38	.70	.92

Computations of the probability of a Type I error in two different situations illustrate how the use of standard hypothesis tests derived under (1) and (2) can be extremely misleading under the autoregressive error model described by (8) and (9).

In Table 4, example 1 is for a two-sided t-test of  $\beta_1 = 0$  in the model:  $y_t = \beta_0 + \beta_1 t + e_t$ ,  $t = 1, \dots, p$  and  $p$  large, Elashoff (1968). Example 2 is for the F-test of  $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$  in the model  $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + e_t$ ,  $t = \frac{1}{25}, \frac{2}{25}, \dots, 1$ , Hoel (1964). Both are examined under the autoregressive error model

$$e_t = \rho e_{t-1} + u_t$$

where  $E(e_t) = 0$  and

$$\Sigma = \sigma_e^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{p-1} \\ \rho & 1 & \rho & \dots & \rho^{p-2} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \rho^{p-1} & \rho^{p-2} & \dots & \dots & 1 \end{bmatrix}.$$

It is apparent that the probability of a Type I error is strongly affected by  $\rho$ . When positive serial correlation exists, the null hypothesis is likely to be rejected far too frequently.

If the ordinary least squares estimators are unbiased and relatively efficient even when serial correlation exists, why are hypothesis tests so misleading? When serial correlation exists, the assumptions of independence of errors which justify the derivation of t- and F- statistics

no longer hold (Elashoff, 1968). In particular, the formula (5) for the variances of estimators no longer holds, and the usual estimator  $\hat{\sigma}_e^2$  of the  $\sigma_e^2$  is a serious underestimate when serial correlation exists.

The general formula for the variance of estimates under assumptions of independence is

$$(15) \quad V(\hat{\beta}) = \sigma_e^2 (X'X)^{-1}.$$

When dependence exists the appropriate formula is

$$(16) \quad V(\hat{\beta}) = (X'X)^{-1} X' \Sigma X (X'X)^{-1}.$$

For example, the standard test of  $H_0: \beta_1 = 0$  is based on using the formula

$$(17) \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma_e^2}{\sum_t (t - \bar{t})^2}$$

for the variance of  $\hat{\beta}_1$  in the formula for the t-statistic. However, if the errors follow the autoregressive model given by (8) and (9) the correct formula for the variance for  $p$  large is approximately

$$(18) \quad \text{Var}(\hat{\beta}_1) \approx \frac{\sigma_e^2}{\sum_t (t - \bar{t})^2} \left(1 + \frac{2\rho}{1-\rho}\right).$$

The difference between these formulae can be quite substantial; their ratio is  $\frac{1-\rho}{1+\rho}$  (see Table 5).

TABLE 5  
Ratio of Standard Formula for  $\text{Var } \hat{\beta}_1$  to Correct Formula for  
 $\text{Var } \hat{\beta}_1$  Under the Autoregressive Error Model

	$\rho$						
	-0.6	-0.4	-0.2	0	0.2	0.4	0.6
$\frac{1-\rho}{1+\rho}$	4.0	2.33	1.5	1	0.75	0.63	0.57

The usual estimator  $\hat{\sigma}_e^2$  of the error variance, obtained by determining the mean square of the residuals from regression, is seriously biased downward. Observation of the graphs in Figure 3 gives some insight into this problem: the least squares lines actually "fit" better than they should, in the sense of reducing the squares of the deviations from the regression line. Hence the residual mean square is much less than the error variance. The average extent of bias under particular polynomial regression models can be determined algebraically. The problem has been investigated by Watson (1955) and Watson and Hannan (1956) at a general theoretical level.

#### Econometric Studies of the Effects of Serial Correlation

In exploring the literature dealing with serial correlation and its effects, much of it will be found to be based on a model appropriate to economic studies (e.g., Cochrane & Orcutt, 1949; Johnston, 1963; Rao & Griliches, 1969). The econometric model is sufficiently different from the polynomial regression model with autoregressive error that the conclusions

based on it are either not directly pertinent to the polynomial model, or need to be interpreted with caution. However, the econometric model has received considerable study and sheds light on the effects of serial correlation.

Econometric models are of the form

$$(19) \quad y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + e_t$$

where  $\beta_0$  is frequently taken to be 0. (In an econometric context, the  $x_{it}$  may be such variables as annual prices of some commodities.) The major difference from the polynomial regression model considered in this paper is the use of random variables  $x_i$  measured at each time point rather than powers of  $t$  to predict  $y$ . More sophisticated treatment of the "time" measure or the use of other indicators of change in conditions could lead to the application of econometric-type models. T. W. Anderson (1963) has pointed out that econometricians generally consider the inclusion of time variables to be substitutes for other unknown variables whose values are related to the time measure.

In particular, Rao and Griliches (1969) study a simple econometric model:

$$(20) \quad \begin{aligned} y_t &= \beta x_t + e_t \\ x_t &= \lambda x_{t-1} + v_t \\ e_t &= \rho e_{t-1} + u_t \end{aligned}$$

$$\begin{aligned}
 E(v_t) &= E(u_t) = E(v_t u_t) \\
 &= E(u_t u_{t-1}) = E(v_t v_{t-1}) = 0
 \end{aligned}$$

$$E(v_t^2) = \sigma_v^2, \quad E(u_t^2) = \sigma_u^2,$$

$$|\lambda| < 1, \quad |\rho| < 1.$$

In contrast, the linear regression model with autoregressive error term is

$$y_t = \beta_0 + \beta_1 t + e_t$$

$$e_t = \rho e_{t-1} + u_t$$

(21)

$$E(u_t) = E(u_t u_{t-1}) = 0$$

$$E(u_t^2) = \sigma_u^2, \quad |\rho| < 1.$$

Examining both models, it will be noted that in model (20) the constant term  $\beta_0$  has been dropped and a random variable  $x_t$  with an autoregressive probability structure has been substituted for  $t$ . A comparison of  $t$  written as  $t = (t-1) + 1$  with  $x_t = \lambda x_{t-1} + v_t$  suggests that results for the Rao and Griliches model for  $\lambda$  close to 1 should be most similar to results for the linear regression model.

For model (21), efficiency computations may be made directly; as seen in Table 3, the ordinary least squares estimators are generally efficient. For econometric models similar to (20), however, efficiency computations based on sampling experiments (necessary because  $x_t$  is a random variable) indicate that ordinary least squares estimators may not be

satisfactory (see Cochrane & Orcutt, 1949, and Rao & Griliches, 1969). Rao and Griliches (1969) indicate that for  $\lambda$  close to 1 ordinary least squares estimators are relatively efficient in the econometric model as one would expect.

### Tests for Serial Correlation

The existence of serial correlation may be investigated in a variety of ways. Figure 3 indicates a common feature of serially correlated data in the tendency for runs of positive and negative residuals from the fitted curve. An obvious initial step is to graph the data and to consider the patterns in the residuals. Mandel (1957, p. 562) applies a test for the cumulative error model based on the number of times the residuals change their sign.

Several tests for the existence of serial correlation of the autoregressive kind have been proposed. Probably the most easily applied is that based on the Durbin-Watson d-statistic, defined as

$$(22) \quad d = \frac{\sum_{t=2}^p (z_t - z_{t-1})^2}{\sum_{t=1}^p z_t^2}$$

where the  $z_t$  are the residuals from the fitted regression line.

Durbin and Watson (1951) provide tables of lower and upper bounds  $d_{\alpha}^L$  and  $d_{\alpha}^U$  for values of  $p$  from 15 to 100 and of  $k$  (degree) from 1 to 5, for single tail significance level  $\alpha = .05$ ,  $.025$ , and  $.001$ . The null hypothesis is  $\rho = 0$  in the model  $e_t = \rho e_{t-1} + u_t$ .

Against the alternative hypothesis  $\rho > 0$ , the null hypothesis is rejected if  $d < d_{l_\alpha}$ , not rejected if  $d > d_{u_\alpha}$ , and the test is inconclusive otherwise. Against the alternative hypothesis  $\rho < 0$ , the null hypothesis is rejected if  $4-d < d_{l_\alpha}$ , not rejected if  $4-d < d_{u_\alpha}$ , and the test is inconclusive otherwise. The alternative hypothesis  $\rho \neq 0$  may be tested by a combination of the above tests at the significance level  $\frac{\alpha}{2}$ . For example, for tests of the hypothesis  $\rho > 0$  for the data of Figure 3, the  $d$  value of 1.98 for the model with  $\rho = 0.2$  is not significant at the .05 level, while the  $d$  value of 0.90 for the model with  $\rho = 0.8$  is significant at the .01 level. Durbin (1970) provides an exact test when this bounds test is inconclusive. For a polynomial regression on  $t$  up to degree 5 Theil and Nagar (1961) provide approximate 1% and 5% significance points.

Durbin (1969) has also developed a graphical method for a more general test of departures from serial independence.

### Conclusions

Studies of regression problems with serial dependence in the residuals have been concerned primarily with evaluating the effects of violating the ordinary least squares assumptions. As pointed out here, ordinary least squares estimators of the coefficients in a polynomial regression such as

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + e_t$$

are unbiased and may still be efficient even when  $\Sigma_e \neq \sigma^2 I$ . However, inferences about  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  based on ordinary least squares procedures may be quite misleading.

Little consideration has been given, in the case of individual series, to the development of alternative procedures or to the problem of providing a more reliable basis for statements of inference, except for the econometric model. Although there are no clear-cut procedures to follow, if it has been determined that there is serial dependence of an extent to make ordinary least squares inappropriate, several alternative possibilities have been suggested.

The initial problem is to settle on an appropriate model for the form of dependence. If the choice is restricted to either a cumulative or a first-order autoregressive model, investigation of the residuals from a fitted ordinary least squares regression may provide sufficient information to distinguish the more appropriate model. Selection of a first-order autoregressive model against a higher-order model may be based on consideration of the sample serial correlation coefficients of lag 1 and higher (Holtzman, 1963). There are some potential problems in arriving at an appropriate model for the form of serial dependence, due to the bias toward randomness of the residuals mentioned in the section concerning the effects of serial correlation on conventional least squares procedures.

Procedures for analyzing data for which a cumulative error model is appropriate are detailed in the articles of Mandel (1959) and Jaech (1964). In a first-order autoregressive model, C. R. Rao (1967) presents a procedure for estimating the coefficients  $\beta$  in a polynomial regression. Closely related to this are the approaches investigated by P. Rao and Griliches (1969).

C. R. Rao (1967), noting that in the first-order autoregressive error model one can write  $u_t = e_t - \rho e_{t-1}$ , expresses the polynomial model in the form

$$u_t = y_t - \rho y_{t-1} - \beta_0(1-\rho) - \beta_1(t-\rho(t-1)) - \dots - \beta_k(t^k - \rho(t-1)^k) .$$

Then  $\rho$  and the  $\beta_i$  are estimated simultaneously by minimizing  $\sum u_t^2$ . This approach looks interesting but no specific information is available on the characteristics of the estimators obtained.

Rao and Griliches have suggested similar approaches for the econometric model (20) although minimizing  $\sum u_t^2$  presents problems in the general situation of random  $x_t$  since the relationship is nonlinear in  $\beta$  and  $\rho$ . Their investigations do suggest that C. R. Rao's approach may be useful for the polynomial regression model.

While consideration of the analysis of individual growth curves provides some insight into the problems presented by the occurrence of serial dependence, educational data is most often available for more than one individual. In this situation, the problem of postulating a model and estimating several parameters, all on a small piece of data, are avoided if one is willing to make certain assumptions about similarities of models for individuals. Gaito and Wiley (1963) and Bock (1963) provide an introduction to the literature in this area.

### References

- Anderson, T. W. The use of factor analysis in the statistical analysis of multiple time series. Psychometrika, 1963, 28, 1-26.
- Anderson, T. W. The statistical analysis of time series. New York: Wiley (in press).
- Bock, R. D. Multivariate analysis of variance of repeated measurements. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.
- Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498.
- Cochrane, D., & Orcutt, G. H. Application of least squares regression to relationships containing auto-correlated error terms. Journal of the American Statistical Association, 1949, 44, 32-61.
- Cronbach, L. J., & Furby, L. How we should measure 'change'--Or should we? Psychological Bulletin, 1970, 74, 68-80.
- Durbin, J. Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. Biometrika, 1969, 56, 1-15.
- Durbin, J. An alternative to the bounds test for testing for serial correlation in least-squares regression. Econometrica, 1970, 30, 422-429.
- Durbin, J., & Watson, G. S. Testing for serial correlation in least-squares regression. I. Biometrika, 1950, 37, 409-428.
- Durbin, J., & Watson, G. S. Testing for serial correlation in least-squares regression. II. Biometrika, 1951, 38, 159-178.
- Elashoff, R. M. Effects of errors in statistical assumptions. In International encyclopedia of the social sciences. New York: Macmillan and Free Press, 1968. Pp. 132-142.
- Estes, W. K. The problems of inference from curves based on group data. Psychological Bulletin, 1956, 53, 134-140.
- Fisher, R. A., & Yates, F. Statistical tables for biological, agricultural and medical research. New York: Hafner, 1957.
- Gaito, J., & Wiley, D. E. Univariate analysis of variance procedures in the measurement of change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.

- Gottman, J. M., McFall, R. M., & Barnett, J. T. Design and analysis of research using time series. Psychological Bulletin, 1969, 72, 299-306.
- Grizzle, J. E., & Allen, D. M. Analysis of growth and dose response curves. Biometrics, 1969, 25, 357-381.
- Hoel, P. G. Methods for comparing growth type curves. Biometrics, 1964, 20, 859-872.
- Holtzman, W. H. Statistical models for the study of change in the single case. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.
- Jaech, J. L. A note on the equivalence of two methods of fitting a straight line through cumulative data. Journal of the American Statistical Association, 1964, 59, 863-866.
- Johnston, J. Econometric methods. New York: McGraw-Hill, 1963.
- Lord, F. M. The measurement of growth. Educational and Psychological Measurement, 1956, 16, 421-437, also, "Errata." Ibid, 1957, 17, 452.
- Lord, F. M. Further problems in the measurement of growth. Educational and Psychological Measurement, 1958, 18, 437-454.
- Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change. Madison: University of Wisconsin Press, 1963.
- Mandel, J. Fitting a straight line to certain types of cumulative data. Journal of the American Statistical Association, 1957, 52, 552-566.
- McGuire, T., & Glass, G. A program for the analysis of certain time-series quasi-experiments. Educational and Psychological Measurement, 1967, 27, 743-750.
- McNemar, Q. On growth measurement. Educational and Psychological Measurement, 1958, 18, 47-55.
- Parzen, E. Mathematical considerations in the estimation of spectra. Technometrics, 1961, 3, 167-190.
- Rao, C. R. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrika, 1965, 52, 447-458.
- Rao, C. R. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, I, 355-372.

- Rao, P., & Griliches, Z. Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. Journal of the American Statistical Association, 1969, 64, 253-272.
- Rosenblatt, M. Some regression problems in time series analysis. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1956, I, 165-186.
- Stake, R. E. Learning parameters, aptitudes, and achievements. Psychometric Monographs, 1961, No. 9.
- Theil, H., & Nagar, A. Testing the independence of regression disturbances. Journal of the American Statistical Association, 1961, 56, 793-806.
- Watson, G. S. Serial correlation in regression analysis. I. Biometrika, 1955, 42, 327-341.
- Watson, G. S., & Hannan, E. J. Serial correlation in regression analysis. II. Biometrika, 1956, 43, 436-448.
- Werts, C. E., & Linn, R. L. A general linear model for studying growth. Psychological Bulletin, 1970, 73, 17-22.
- Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

## APPENDIX

## Orthogonal Polynomial Coefficients

This paper discusses the effects of serial dependence on analyses based on polynomial regression coefficients. Generally, orthogonal polynomial coefficients have been preferred in the psychological and educational literature dealing with repeated measurements data as being more descriptive of the data (see, e.g., Gaito & Wiley, 1963; Bock, 1963). Some of the relationships between the two approaches will be briefly indicated here.

In equation (3) the general form of the polynomial regression model is expressed as

$$(A.1) \quad \underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{e}$$

where

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 4 & \dots & 2^{k-1} \\ 1 & 3 & 9 & \dots & 3^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & p & p^2 & \dots & p^{k-1} \end{bmatrix}$$

for equally spaced time points.

Analysis based on orthogonal polynomial coefficients proceeds in the same way as for the polynomial regression, but the matrix  $\underset{\sim}{X}$  is replaced by a matrix of orthogonal polynomials, denoted for convenience as  $\underset{\sim}{P}$ . The vector of coefficients is denoted as  $\underset{\sim}{\gamma}$ . Hence the model (3) may be expressed as

$$(A.2) \quad \underset{\sim}{y} = \underset{\sim}{P} \underset{\sim}{\gamma} + \underset{\sim}{e}.$$

The matrix  $\underset{\sim}{P}$  is obtained by a process of "orthogonalizing" the matrix  $\underset{\sim}{X}$ . Standard procedures exist for this, but the elements of  $\underset{\sim}{P}$  are readily available in tabulated form when the time points are equally spaced (Winer, 1962; Fisher & Yates, 1957). For unequally spaced time points, see Gaito (1965).

As an example, the matrix  $\underset{\sim}{P}$  corresponding to a quadratic polynomial based on 5 time points is

$$\underset{\sim}{P} = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{10}} & \frac{2}{\sqrt{14}} \\ \frac{1}{\sqrt{5}} & \frac{-1}{\sqrt{10}} & \frac{-1}{\sqrt{14}} \\ \frac{1}{\sqrt{5}} & 0 & \frac{-2}{\sqrt{14}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{10}} & \frac{-1}{\sqrt{14}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{10}} & \frac{2}{\sqrt{14}} \end{bmatrix} .$$

$\underset{\sim}{P}$  has the characteristic that the columns are orthogonal to each other; i.e., the sum of the products of the elements is zero. This leads to the result:

$$\underset{\sim}{P}'\underset{\sim}{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \underset{\sim}{I} .$$

Consequently, the formulae for  $\hat{\gamma}$  and  $\text{Var } \hat{\gamma}$  are considerably simpler than for  $\hat{\beta}$  and  $\text{Var } \hat{\beta}$ . The estimators of  $\gamma$  from least squares theory are:

$$(A.3) \quad \hat{\gamma}_{\sim} = P' y_{\sim}$$

with variance-covariance matrix

$$(A.4) \quad V(\hat{\gamma}_{\sim}) = \sigma^2 I.$$

This latter formula indicates the algebraic independence among the estimates of the coefficients when the errors are independent. Since the variance-covariance matrix is diagonal, the individual coefficients are independent; e.g., in the example shown in Table 2,  $\hat{\gamma}_0$ ,  $\hat{\gamma}_1$ ,  $\hat{\gamma}_2$  are algebraically independent. This is not true for the polynomial regression coefficients, for which there exists an algebraic dependence between  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\beta}_2$ , etc.

When serial dependence exists between observations, formula (14) for the variance of the ordinary least squares estimators becomes

$$\begin{aligned} \text{Var } (\hat{\gamma}_{\sim}) &= (P'P)^{-1} P' \Sigma P (P'P)^{-1} \\ &= P' \Sigma P \quad \text{since } P'P = I. \end{aligned}$$

It can be seen from this formula that when  $\Sigma \neq \sigma^2 I$  the o.l.s. estimates of the orthogonal polynomial coefficients no longer have the advantage of being independent. Relative efficiency figures derived from these formulae for the orthogonal polynomial coefficients are quite similar to those for the ordinary polynomial coefficients, and hypothesis testing behavior should follow the same pattern.